539,890

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 July 2004 (15.07.2004)

PCT

(10) International Publication Number
WO 2004/059536 A2

(51) International Patent Classification⁷: G06F 17/30, 3/033

(21) International Application Number: PCT/US2003/040726

(22) International Filing Date: 19 December 2003 (19.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/325,061 20 December 2002 (20.12.2002) US

(71) Applicant *(for all designated States except US)*: **INTERNATIONAL BUSINESS MACHINES COPORATION** [US/US]; Armonk, NY 10504 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ADAMS, Hugh, W., Jr.** [US/US]; 17 Rolling Green lane, Wappingers Falls, NY 12590 (US). **IYENGAR, Giridharen** [IN/US]; 80 Carey Street, Mahopac, NY 10541 (US). **LIN, Ching-Yung** [CN/US]; 68-43 Nansen Street, Forest Hills, NY 11375 (US). **NETI, Chalapathy, V.** [IN/US]; 235 High Ridge Court, Yorktown Heights, NY 10598 (US). **SMITH, John, R.** [US/US]; 40 Farrel Street, New Hyde Park, NY 11040 (US). **TSENG, Belle, L.** [US/US]; 68-49 Nanse Street, Forest Hills, NY 11375 (US).

(74) Agent: **FERENCE, Stanley, D., III.**; Ference & Associated, 400 Broad Street, Pittsbury, PA 15143 (US).

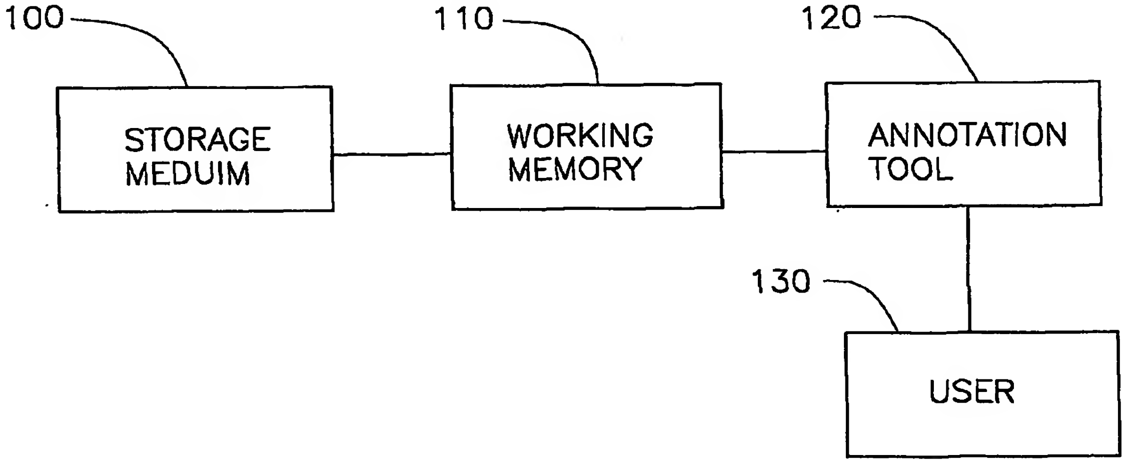
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR ANNOTATING MULTI-MODAL CHARACTERISTICS IN MULTIMEDIA DOCUMENTS



(57) Abstract: A manual annotation system of multi-modal characteristics in multimedia files. There is provided an arrangement for selection an observation modality of video with audio, video without audio, audio with video, or audio without video, to be used to annotate multimedia content. While annotating video or audio features is isolation results in less confidence in the identification of features, observing both audio and video simultaneously and annotating that observation results in a higher confidence level.

SYSTEM AND METHOD FOR ANNOTATING MULTI-MODAL CHARACTERISTICS IN MULTIMEDIA DOCUMENTS

Field of the Invention

The present invention relates to the computer processing of multimedia files. More specifically, the present invention relates to the manual annotation of multi-modal events, objects, scenes, and audio occurring in multimedia files.

Background of the Invention

Multimedia content is becoming more common both on the World Wide Web and local computers. As the corpus of multimedia content increases, the indexing of features within the content becomes more and more important. Observing both audio and video simultaneously and annotating that observation results in a higher confidence level.

Existing multimedia tools provide capabilities to annotate either audio or video separately, but not as a whole. (An example of a video-only annotation tool is the IBM MPEG7 Annotation Tool, inventors J. Smith et al., available through [http://]www.alphaworks.ibm.com/tech/videoannex. Other conventional arrangements are described in: Park et al, "iMEDIA-CAT: Intelligent Media Content Annotation Tool", Proc. International Conference on Inductive Modeling (ICIM)

2001, South Korea, November, 2001; and Minka et al., "Interactive Learning using a Society of Models," Pattern Recognition, Vol. 30, pp. 565, 1997, TR #349.

It has long been recognized that annotating video or audio features in isolation results in a less confidence of the identification of the features.

In view of the foregoing, a need has been recognized in connection with providing improved systems and methods for observing and annotating multi-modal events, objects, scenes, and audio occurring in multimedia files.

Summary of the Invention

In accordance with at least one presently preferred embodiment of the present invention, there are broadly contemplated multimedia annotation systems and methods that permit users to observe solely video, video with audio, solely audio, or audio with video and to annotate what has been observed.

In one embodiment, there is provided a computer system which has one or more multimedia files that are stored in a working memory. The multi-modal annotation process displays a user selected multimedia file, permits the selection of a mode or modes to observe the file content, annotates the observations; and saves the annotations in a working memory (such as a MPEG-7 XML file).

In summary, one aspect of the invention provides an apparatus for managing multimedia content, the apparatus comprising: an arrangement for supplying multimedia content; an input interface for permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and an arrangement for annotating observations of a selected mode.

A further aspect of the invention provides a method of managing multimedia content, the method comprising the steps of: supplying multimedia content; permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and annotating observations of a selected mode.

Furthermore, an additional aspect of the invention provides a program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for managing multimedia content, the method comprising the steps of: supplying multimedia content; permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and annotating observations of a selected mode.

For a better understanding of the present invention, together with other and further features and advantages thereof, reference is made to the following description, taken in conjunction with the accompanying drawings, and the scope of the invention will be pointed out in the appended claims.

Brief Description of the Drawings

Figure 1 is a block diagram depicting a multi-modal annotation system.

Figure 2 is an illustration of a system annotating video scenes, objects, and events.

Figure 3 is an illustration of a system annotating audio with video.

Figure 4 is an illustration of a system annotating audio without video.

Description of the Preferred Embodiments

Figure 1 is a block diagram of one preferred embodiment of a multi-modal annotation system in accordance with the present invention. The multimedia content and previous annotations are stored on the storage medium 100. When a user 130 selects a multimedia file via the annotation tool from the storage medium 100, it is loaded into working memory 110 and portions of it displayed in the annotation tool 120. At any time, the user 130 may also request that previously saved annotations associated with the current multi-modal file be loaded from the storage medium 100 into working memory 110. The user 100 views the multimedia data by making requests through the annotation tool 120. The user 130 then annotates his observations and the annotation tool 120 saves these annotations in working memory 110. The user can at anytime request the annotation tool 120 to save the annotation on the storage medium 100.

Figure 2 is an illustration of a system annotating video scenes, objects, and events. (Simultaneous reference should also be made to Fig. 1.) The multimedia data has been loaded from the storage medium 100 into working memory 110. A video tab 290 has been selected. The multimedia video has been segmented using scene changed detection into shots. A shot list window 200 displays a portion of the shots in the multimedia. Here, the user 130 has selected a shot 210 which is highlighted in the shot list window 200. A key frame 220, which is a representative shot in the frames of a shot, is preferably displayed. In addition, the frames of the shot maybe viewed in the video window 230 using play controls 240. The video can be viewed with or without audio depending upon the selection of a mute button 250. The user 130 may select annotations for this shot by clicking the boxes in events 260, static scenes 270, or key objects 280 lists of boxes. Any significant observations which are not contained in the check boxes can be noted in a keywords text box 300.

Figure 3 is an illustration of the system annotating audio with video. (Simultaneous reference should also be made to Fig. 1.) The multimedia data has been loaded from the storage medium 100 into working memory 110. The audio with video tab 370 has been selected. The multimedia video has been segmented using scene change detection into shots. The shot list window 200 displays a portion of the shots in the multimedia. The shot 210 associated with the current audio position is highlighted in the shot list window 200. The audio data is displayed in the window 390. A segment of audio 340 has been delimited for annotation; that is, the limits or bounds of the audio has been fixed for subsequent annotation. The video associated with the audio is shown in 230. As the user 130 uses the play controls 360, the audio data display 390 is updated to display the current audio data and the video window 230 changes to reflect the current video frame. Thus, the user 130 may observe the video and simultaneously hear the audio while making audio annotations. The user 130 preferably uses the buttons 350 to delimit audio segments. Check boxes corresponding to the foreground sounds (320) (the most prominent sounds in the segment) and background sounds (330) (sounds which are present but are secondary to other sounds) may be checked to indicated sounds heard within the audio segment 340. Any significant observations which are not contained in the check boxes can be noted in keywords text box 300.

Figure 4 is an illustration of the system annotating audio without video. (Simultaneous reference should be made to Fig. 1.) The multimedia data has been loaded from the storage medium 100 into working memory 110. Audio-without-video tab 400 has been selected. The audio data is displayed in the window 390. A segment of audio 340 has been delimited for annotation. As the user 130 uses the play controls 360, the audio data display 390 is updated to display the current audio data. Thus, the user 130 may only hear the audio while making audio annotations. The user 130 uses the buttons 350 to delimit audio segments. The check boxes for foreground sounds 320 and background sounds 330 may be checked to indicate sounds heard within the audio segment 340. Any significant observations which are not contained in the check boxes can be noted in the keywords text box 300.

It is to be understood that the present invention, in accordance with at least one presently preferred embodiment, includes an arrangement for supplying multimedia content, an input interface for permitting the selection, for observation, of a mode associated with the multimedia content, and an arrangement for annotating observations of a selected mode. Together, these elements may be implemented on at least one general-purpose computer running suitable software programs. These may also be implemented on at least one Integrated Circuit or part of at least one Integrated Circuit. Thus, it is to be understood that the invention may be implemented in hardware, software, or a combination of both.

If not otherwise stated herein, it is to be assumed that all patents, patent applications, patent publications and other publications (including web-based publications) mentioned and cited herein are hereby fully incorporated by reference herein as if set forth in their entirety herein.

Although illustrative embodiments of the present invention have been described herein with reference to the accompanying drawings, it is to be understood that the invention is not limited to those precise embodiments, and that various other changes and modifications may be affected therein by one skilled in the art without departing from the scope or spirit of the invention.

Claims

What is claimed is:

1. An apparatus for managing multimedia content, said apparatus comprising:

an arrangement for supplying multimedia content;

an input interface for permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and

an arrangement for annotating observations of a selected mode.

2. The apparatus according to Claim 1, wherein said input interface permits the selection, for observation, of both of the following associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio.

3. The apparatus according to Claim 1, wherein said input interface additionally permits the selection, for observation, of solely a video portion of multimedia content.

4. The apparatus according to Claim 1, wherein said input interface additionally permits the selection, for observation, of solely an audio portion of multimedia content.

5. The apparatus according to Claim 1, wherein said arrangement for supplying multimedia content comprises a working memory which stores multimedia files.

6. The apparatus according to Claim 1, wherein said input interface is adapted to: first permit the selection of a multimedia file and then permit the selection of said at least one of: an audio portion simultaneously with video; and a video portion simultaneously with audio.

7. The apparatus according to Claim 1, further comprising a working memory for saving the annotated observations of a selected mode.

8. The apparatus according to Claim 1, wherein said input interface is adapted to permit the selection, for observation, at least the following mode associated with the multimedia content: a video portion that includes audio.

9. The apparatus according to Claim 8, wherein said input interface comprises:

an arrangement for permitting the selection, for observation, of a video mode of multimedia content; and

an arrangement for selectably adding audio to the video mode for observation.

10. A method of managing multimedia content, said method comprising the steps of:

supplying multimedia content;

permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and

annotating observations of a selected mode.

11. The method according to Claim 10, wherein said step of permitting selection comprises permitting the selection, for observation, of both of the following associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio.

12. The method according to Claim 10, wherein said step of permitting selection additionally comprises permitting the selection the selection, for observation, of solely a video portion of multimedia content.

13. The method according to Claim 10, wherein step of permitting selection comprises permitting the selection, for observation, of solely an audio portion of multimedia content.

14. The method according to Claim 10, wherein said step of supplying multimedia content comprises providing a working memory which stores multimedia files.

15. The method according to Claim 10, wherein said step of permitting selection comprises: first permitting the selection of a multimedia file and then permitting the selection of said at least one of: an audio portion simultaneously with video; and a video portion simultaneously with audio.

16. The method according to Claim 10, further comprising the step of providing a working memory for saving the annotated observations of a selected mode.

17. The method according to Claim 10, wherein said step of permitting selection comprises permitting the selection, for observation, at least the following mode associated with the multimedia content: a video portion that includes audio.

18. The method according to Claim 17, wherein said step of permitting selection comprises:

permitting the selection, for observation, of a video mode of multimedia content; and

thereafter enabling the addition of audio to the video mode for observation.

19. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for managing multimedia content, said method comprising the steps of:

supplying multimedia content;

permitting the selection, for observation, of at least one of the following modes associated with the multimedia content: an audio portion that includes video; and a video portion that includes audio; and

annotating observations of a selected mode.

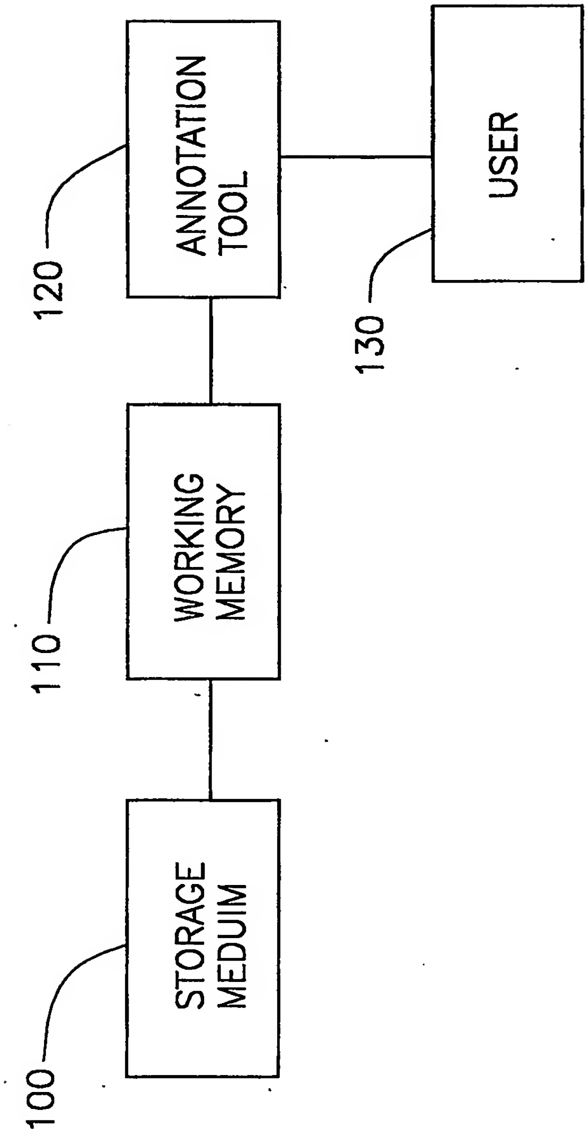
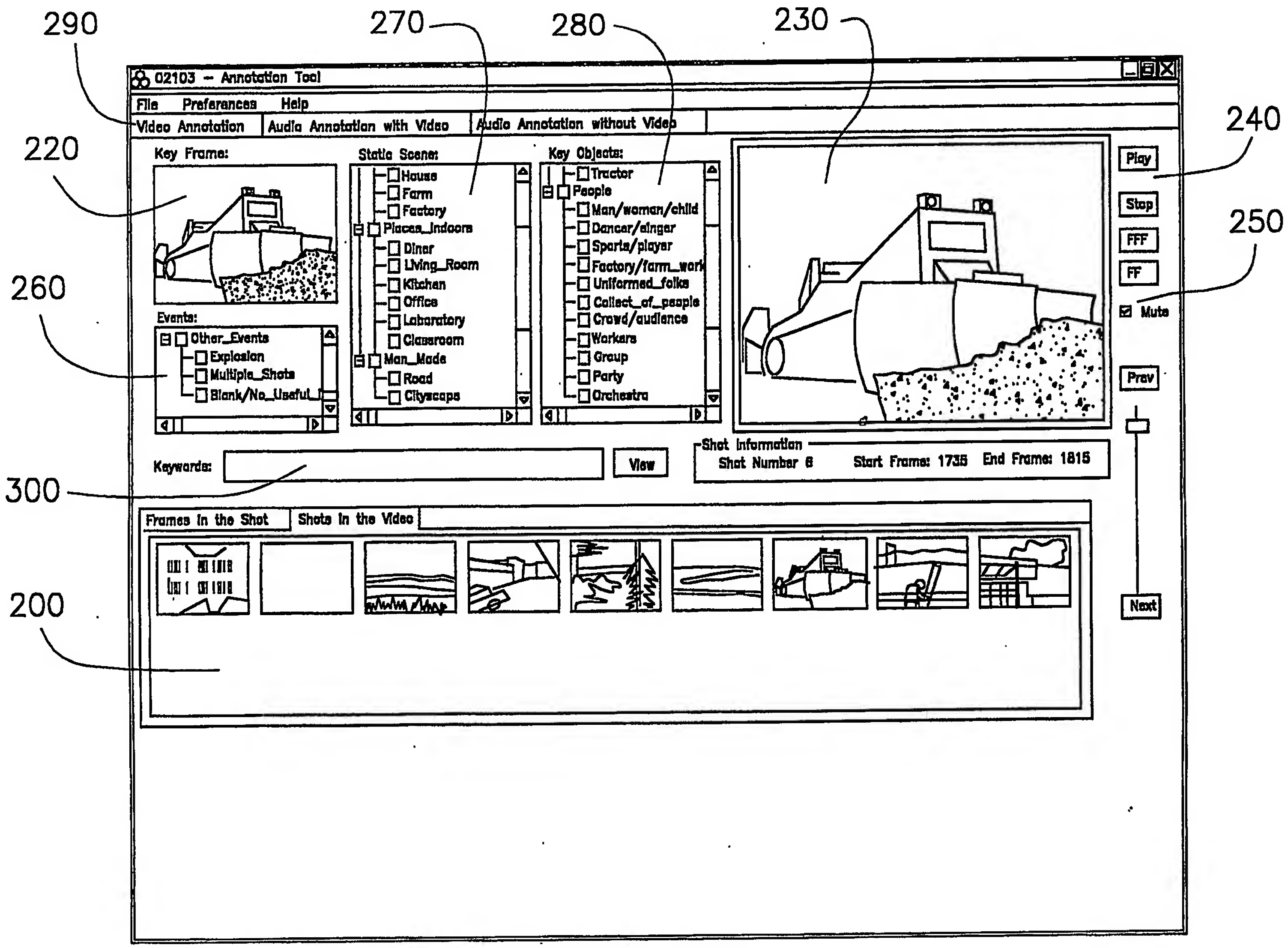
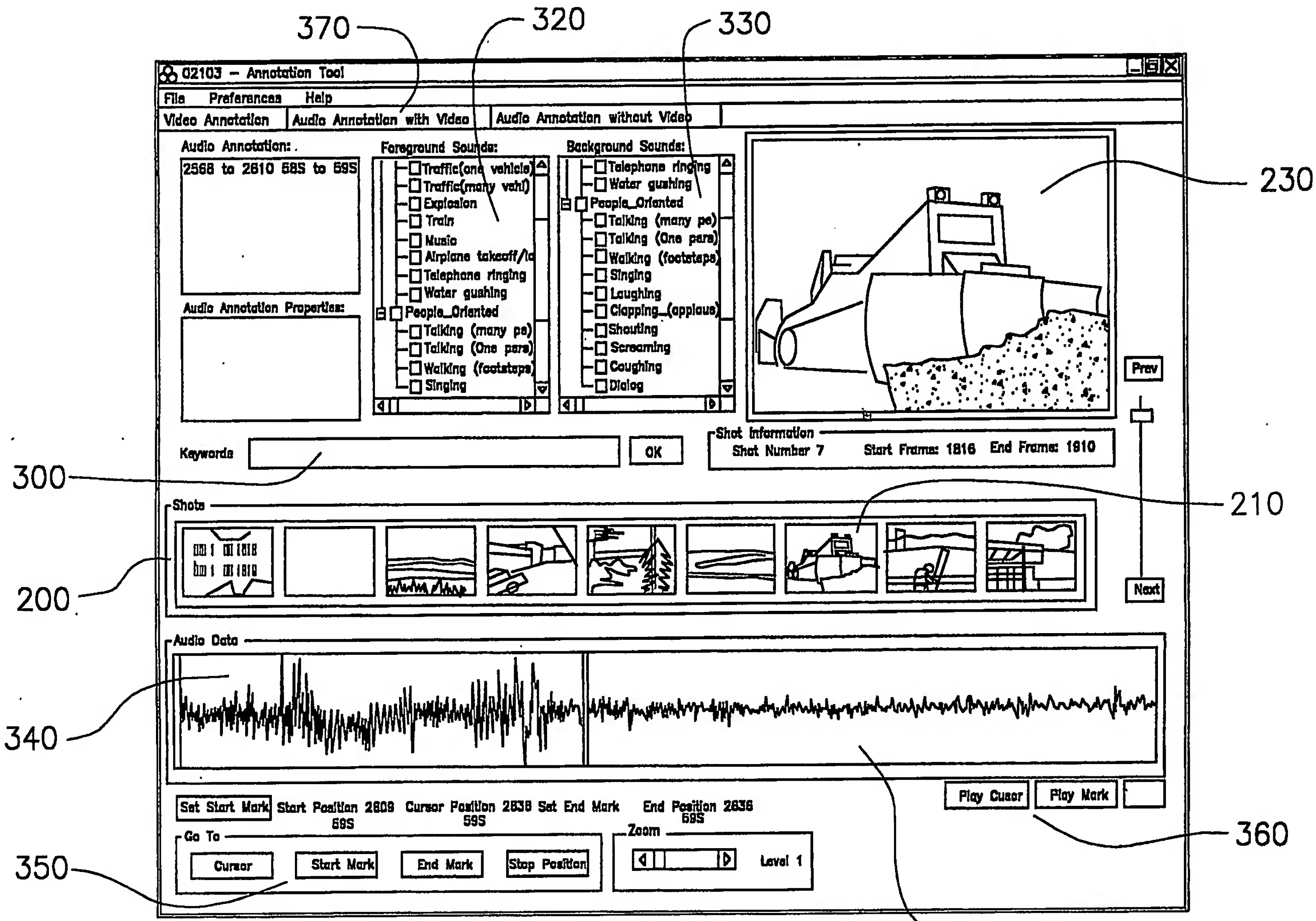

STORAGE MEDUIM
100
WORKING MEMORY
110
ANNOTATION TOOL
120
USER
130

FIG. 1



Shot Information
Shot Number 8 Start Frame: 1735 End Frame: 1815

FIG. 2



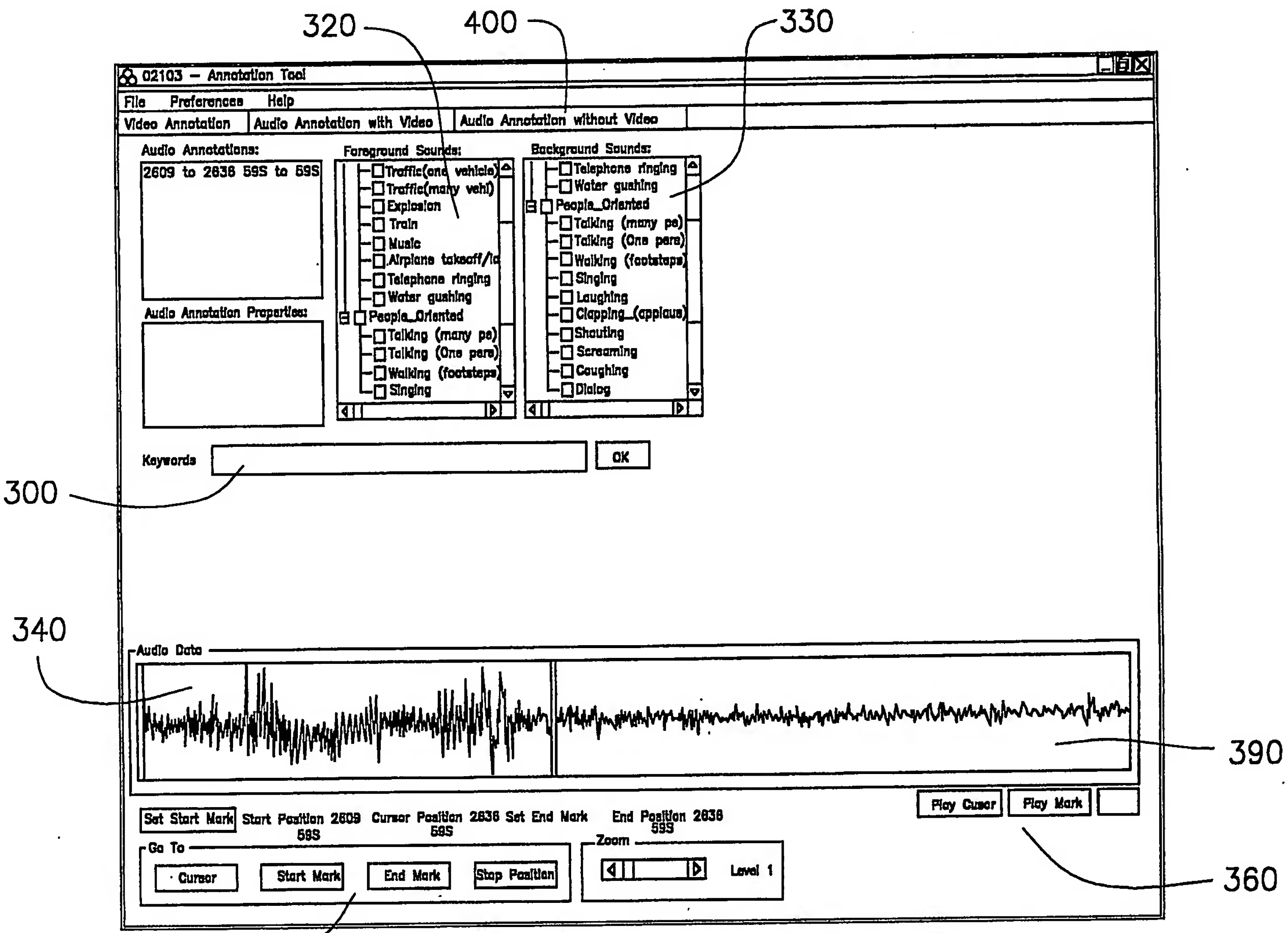
2568 to 2610 58S to 59S

Shot Number 7 Start Frame: 1816 End Frame: 1910

Start Position 2609 59S Cursor Position 2638 59S End Position 2636 59S

Level 1

FIG. 3



02103 – Annotation Tool

Audio Annotations:
2609 to 2636 59S to 59S

Start Position 2609 59S
Cursor Position 2636 59S
End Position 2636 59S

Zoom Level 1

FIG. 4

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 July 2004 (15.07.2004)

PCT

(10) International Publication Number
WO 2004/059536 A3

(51) International Patent Classification⁷: G06F 17/30, 3/033

(21) International Application Number: PCT/US2003/040726

(22) International Filing Date: 19 December 2003 (19.12.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/325,061 20 December 2002 (20.12.2002) US

(71) Applicant (*for all designated States except US*): INTERNATIONAL BUSINESS MACHINES COPORATION [US/US]; Armonk, NY 10504 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): **ADAMS, Hugh, W., Jr.** [US/US]; 17 Rolling Green lane, Wappingers Falls, NY 12590 (US). **IYENGAR, Giridharen** [IN/US]; 80 Carey Street, Mahopac, NY 10541 (US). **LIN, Ching-Yung** [CN/US]; 68-43 Nansen Street, Forest Hills, NY 11375 (US). **NETI, Chalapathy, V.** [IN/US]; 235 High Ridge Court, Yorktown Heights, NY 10598 (US). **SMITH, John, R.** [US/US]; 40 Farrel Street, New Hyde Park, NY 11040 (US). **TSENG, Belle, L.** [US/US]; 68-49 Nanse Street, Forest Hills, NY 11375 (US).

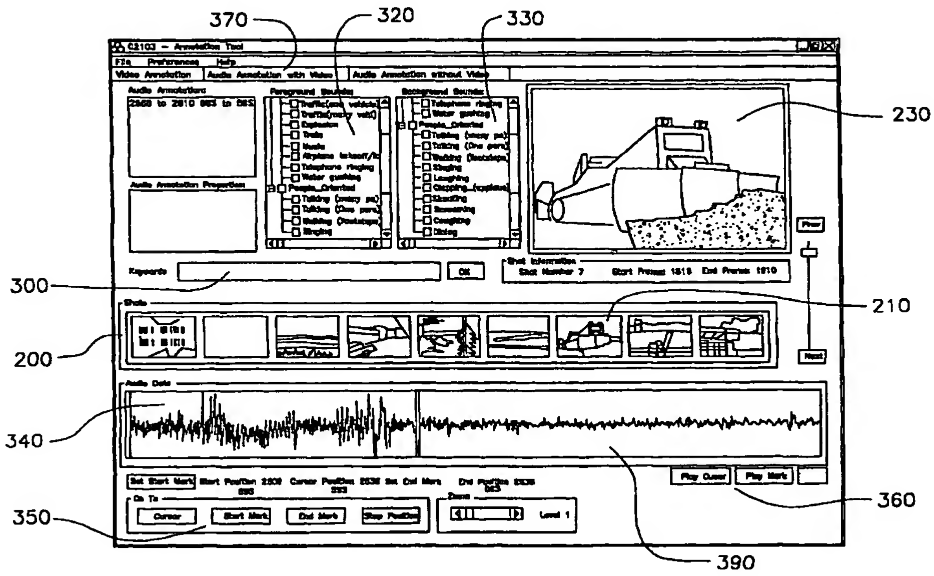
(74) Agent: **FERENCE, Stanley, D., III.**; Ference & Associated, 400 Broad Street, Pittsbury, PA 15143 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR ANNOTATING MULTI-MODAL CHARACTERISTICS IN MULTIMEDIA DOCUMENTS



(57) Abstract: A manual annotation system of multi-modal characteristics in multimedia files. There is provided an arrangement for selection an observation modality of video with audio, video without audio, audio with video, or audio without video, to be used to annotate multimedia content. While annotating video or audio features is isolation results in less confidence in the identification of features, observing both audio and video simultaneously and annotating that observation results in a higher confidence level.

WO 2004/059536 A3

Published:
— *with international search report*

(88) Date of publication of the international search report: 29 December 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 03/40726

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F17/30 G06F3/033

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 442 744 A (MORRIS TREVOR ET AL) 15 August 1995 (1995-08-15) column 4, paragraph 1; claim 1; figure 3 -----	1-19
A	US 2001/013068 A1 (CHOU PHILIP A ET AL) 9 August 2001 (2001-08-09) claim 1; figure 7 -----	1-19
A	US 5 600 775 A (KING PHILIP S ET AL) 4 February 1997 (1997-02-04) claim 1; figure 2 -----	1-19
A	EP 1 158 795 A (SHARP KK) 28 November 2001 (2001-11-28) claim 1; figure 1 ----- -/--	1-19

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search 4 October 2004	Date of mailing of the international search report 11/10/2004
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer Kirsten, K

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 03/40726

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 489 576 A (SONY CORP AMERICA) 10 June 1992 (1992-06-10) claim 1; figures 4,6 -----	1-19
A	TSENG B L ET AL: "Video personalization and summarization system" IEEE, 9 December 2002 (2002-12-09), pages 424-427, XP010642599 figure 2 -----	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 03/40726

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
US 5442744	A	15-08-1995	JP	6043839	A	18-02-1994
US 2001013068	A1	09-08-2001	NONE			
US 5600775	A	04-02-1997	NONE			
EP 1158795	A	28-11-2001	EP	1158795	A2	28-11-2001
			JP	2002077786	A	15-03-2002
EP 0489576	A	10-06-1992	US	5148154	A	15-09-1992
			DE	69131065	D1	06-05-1999
			DE	69131065	T2	04-11-1999
			DE	69132981	D1	08-05-2002
			DE	69132981	T2	28-11-2002
			DE	69133127	D1	07-11-2002
			DE	69133127	T2	05-06-2003
			EP	0489576	A2	10-06-1992
			EP	0786716	A2	30-07-1997
			EP	0786717	A2	30-07-1997
			JP	3165815	B2	14-05-2001
			JP	4307675	A	29-10-1992
			US	5307456	A	26-04-1994